Supplementary Documents for “Semi-Crowdsource Clustering: Generalizing Crowd Labeling by Robust Distance Metric Learning”

Jinfeng Yi†, Rong Jin†, Anil K. Jain†, Shaili Jain♮, Tianbao Yang‡
†Michigan State University, East Lansing, MI 48824, USA
♮Yale University, New Haven, CT 06520, USA
‡Machine Learning Lab, GE Global Research, San Ramon, CA 94583, USA
{yijinfen, rongjin, jain}@cse.msu.edu, shaili.jain@yale.edu, tyang@ge.com

1 Theoretical Analysis for Perfect Recovery using Equation (2)

The following discussion about the perfect recovery result using Eq. (2) comes from [3]. We repeat it in the supplementary document for the completeness of this study.

To discuss the perfect recovery result for using Eq. (2), we first need to make a few assumptions about $A^*$ besides its low rank. Let $A^*$ be a low-rank matrix of rank $r$, with a singular value decomposition $A^* = U\Sigma V^T$, where $U = (u_1, \ldots, u_r) \in \mathbb{R}^{N \times r}$ and $V = (v_1, \ldots, v_r) \in \mathbb{R}^{N \times r}$ are the left and right eigenvectors of $A^*$, satisfying the following incoherence assumptions.

- **A1** The row and column spaces of $A^*$ have coherence bounded above by some positive number $\mu_0$, i.e.,
  \[
  \max_{i \in [N]} \|P_U(e_i)\|_2^2 \leq \frac{\mu_0 r}{N}, \quad \max_{i \in [N]} \|P_V(e_i)\|_2^2 \leq \frac{\mu_0 r}{N}
  \]
  where $e_i$ is the standard basis vector.

- **A2** The matrix $E = UV^T$ has a maximum entry bounded by $\frac{\mu_1 \sqrt{r}}{N}$ in absolute value for some positive $\mu_1$, i.e. $|E_{i,j}| \leq \frac{\mu_1 \sqrt{r}}{N}$, $\forall (i,j) \in [N] \times [N]$,

where $P_U$ and $P_V$ denote the orthogonal projections on the column space and row space of $A^*$, respectively, i.e.

\[
P_U = UU^T, \quad P_V = VV^T
\]

To state our theorem, we need to introduce a few notations. Let $\xi(A')$ and $\mu(A')$ denote the low-rank and sparsity incoherence of matrix $A'$ defined by [1], i.e.

\[
\xi(A') = \max_{E \in T(A'), \|E\| \leq 1} \|E\|_\infty \quad (1)
\]

\[
\mu(A') = \max_{E \in \Omega(A'), \|E\|_\infty \leq 1} \|E\| \quad (2)
\]

where $T(A')$ denotes the space spanned by the elements of the form $u_k y^T$ and $x v_k^T$, for $1 \leq k \leq r$, $\Omega(A')$ denotes the space of matrices that have the same support to $A'$, $\| \cdot \|$ denotes the spectral norm and $\| \cdot \|_\infty$ denotes the largest entry in magnitude.

**Lemma 1.** Let $A^* \in \mathbb{R}^{N \times N}$ be a similarity matrix of rank $r$ obeying the incoherence properties (A1) and (A2), with $\mu = \max(\mu_0, \mu_1)$. Suppose we observe $m_1$ entries of $A^*$ recorded in $\tilde{A}$...
Lemma 2. (Lemma 2 from [2]) Let $\mathcal{H}$ be a Hilbert space and $\xi$ be a random variable on $(Z, \rho)$ with values in $\mathcal{H}$. Assume $\|\xi\| \leq M < \infty$ almost surely. Denote $\sigma^2(\xi) = \mathbb{E}([\|\xi\|]^2)$. Let $\{z_i\}_{i=1}^m$ be independent random drawers of $\xi$. For any $0 < \delta < 1$, with confidence $1 - \delta$, 
\[
\left\| \frac{1}{m} \sum_{i=1}^{m} (z_i - \mathbb{E}[z_i]) \right\| \leq \frac{4M \ln(2/\delta)}{\sqrt{m}}
\]

Using the assumption that $|x|_2 \leq 1$ and Lemma 2 we have, with a probability $1 - n^{-3}$, 
\[
\left| \frac{1}{m} \hat{X} \hat{X}^T - C_X \right|_2 \leq \frac{12 \ln n}{\sqrt{n}}
\]

and therefore 
\[
\left| \left( \frac{1}{m} \hat{X} \hat{X}^T + \lambda I \right)^{-1} - (C_X + \lambda I)^{-1} \right|_2 \leq \frac{12 \ln n}{\lambda \sqrt{n}}
\]

Second, according to Lemma 1 with a probability $1 - n^{-3}$, we have $\hat{A} = YY^T$ and therefore $\hat{X}^T \hat{X}^\top \hat{X} = \hat{X} Y Y^T \hat{X}$. Again, using the matrix concentration theory, we have, with a probability $1 - n^{-3}$, 
\[
\left| \frac{1}{m} \hat{X} Y - B \right|_2 \leq \frac{12 \ln n}{\sqrt{n}}
\]

Finally, we rewrite $|M_s - \hat{M}_s|_2$ as 
\[
\begin{align*}
|M_s - \hat{M}_s|_2 & \leq |M_s - \left( \frac{1}{m} \hat{X} \hat{X}^T + \lambda I \right)^{-1} BB^T C_X|_2 + \\
& \leq \left| \left( \frac{1}{m} \hat{X} \hat{X}^T + \lambda I \right)^{-1} BB^T C_X - \left( \frac{1}{m} \hat{X} \hat{X}^T + \lambda I \right)^{-1} \right|_2 + \\
& \leq \left| \left( \frac{1}{m} \hat{X} \hat{X}^T + \lambda I \right)^{-1} BB^T \left( \frac{1}{m} \hat{X} \hat{X}^T + \lambda I \right)^{-1} \hat{X} Y m^{-1} B^T \left( \frac{1}{m} \hat{X} \hat{X}^T + \lambda I \right)^{-1} - \hat{M}_s \right|_2
\end{align*}
\]

It is easy to see that with a probability $1 - 3n^{-3}$, each term on the right hand side of the above inequality is bounded by $\frac{12 \ln n}{\lambda \sqrt{n}}$, leading to the result of the theorem.
References

